Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 2027–2031, March 1992
Biochemistry

YMXM motifs of IRS-1 define substrate specificity of the insulin receptor kinase

(signal transduction/phosphatidylinositol 3′-kinase/oncogene/tyrosine kinase/*src* homology domain 2)

STEVEN E. SHOELSON†, SWATI CHATTERJEE, MANAS CHAUDHURI, AND MORRIS F. WHITE

Research Division, Joslin Diabetes Center, Departments of Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02215

Communicated by Josef Fried, December 6, 1991

ABSTRACT **Of 34 tyrosine residues in insulin receptor substrate 1 (IRS-1), 14 are adjacent to acidic residues, suggesting that they might be phosphorylation sites. Synthetic peptides corresponding to sequences surrounding these tyrosines were used as substrates of the insulin receptor kinase. Surprisingly six of these, each within YMXM motifs, were phosphorylated with greatest efficiency (K_m, 24–92 μM; k_{cat}/K_m, 0.6–2.1 $\times$ 10^4 M^{-1}·sec^{-1}). Substituted YMXM peptides revealed a strong preference of the insulin receptor kinase for methionine at Y + 1 and Y + 3 positions. When phosphorylated, related YMXM sequences are recognition motifs for binding to proteins with *src*-homology (SH2) domains. The combined hydrophobic and flexible nature of methionine side chains adjacent to the targeted tyrosines provides a versatile contact for recognition by diverse proteins involved in signal transduction.**

Insulin binding to the extracellular α subunits of the insulin receptor activates tyrosine kinase activity intrinsic to the intracellular β subunits (1–4). This finding sparked a search for cellular substrates that could link the insulin receptor to postreceptor signaling events. The first identified endogenous substrate of the insulin receptor was pp185, a cytoplasmic phosphoprotein of 165–185 kDa (5). Recently, a cDNA corresponding to a component of the pp185 band was sequenced to provide the deduced primary structure of insulin receptor substrate 1 (IRS-1) (6, 7). As IRS-1 has little extended sequence homology with other known proteins, an intrinsic function for it has not been assigned. IRS-1 is phosphorylated on tyrosine residues after insulin stimulation (7), although the exact sites of phosphorylation have yet to be determined. Of a total of 34 tyrosines, 14 are preceded by acidic amino acids, suggesting that they might be targets of the insulin receptor kinase. Surprisingly, 6 of these are in YMXM motifs and two more are in YXXM motifs.

The platelet-derived growth factor (PDGF) receptor and polyoma virus middle-sized tumor antigen (middle T antigen) contain similar YMXM and homologous YVXM motifs, which are phosphorylated and thought to be essential for binding to phosphatidylinositol 3′-kinase (PI 3-kinase) (8, 9). The 85-kDa regulatory subunit of PI 3-kinase contains two SH2 domains (10, 11), which are thought to mediate its interactions with the PDGF receptor, the $pp60^{c\text{-}src}$/middle T antigen complex, and other proteins involved in cellular signaling (8). In fact, immunoprecipitation studies suggest that PI 3-kinase activity is closely associated with IRS-1 after insulin stimulation of intact cells (7, 12, 13). Thus, phosphorylated IRS-1 might act as a "docking protein" to bind and regulate PI 3-kinase and additional signal-transducing proteins containing SH2 domains as well.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

In this study, we show that peptides corresponding to IRS-1 YMXM motifs are excellent substrates of the insulin receptor kinase. In fact, within a given sequence the presence of methionine residues at Y + 1 and Y + 3 positions is even more important than acidic residues N-terminal to tyrosine for catalytic efficiency. These findings suggest that in addition to being recognition elements for interactions with SH2 domain-containing proteins, residues within YMXM sequences form a recognition motif for insulin receptor-catalyzed phosphorylation. These studies define structural requisites for efficient phosphorylation of a cellular substrate of the insulin receptor and may extrapolate directly to other members of the family of tyrosine kinases and their interactions with additional elements of signaling cascades.

EXPERIMENTAL PROCEDURES

Peptide Synthesis. Solid-phase syntheses were performed on an Applied Biosystems model 430A synthesizer using standard dicyclohexylcarbodiimide-mediated preformed symmetrical anhydride coupling protocols. Amino acids with standard butoxycarbonyl/benzyl protecting groups were purchased from Applied Biosystems; additional solvents and reagents were the highest purity available. Peptide products were cleaved from the resin, side-chain protecting groups were removed, and methionine sulfoxide was reduced during "low-high" cleavages with trifluoromethanesulfonic acid (14). Peptides were typically quite pure; those having ≤95% purity by analytical reversed-phase HPLC were purified further by preparative HPLC (Waters Prep 4000) on a Dynamax-300A 12-μm C8 column (41.4 × 250 mm) equipped with a matched guard column. Peptides were eluted with a mobile phase composed of acetonitrile in 0.05% aqueous trifluoroacetic acid (80 ml/min). Amino acid analyses of all peptides used for kinetic analyses were as predicted.

Insulin Receptor Preparation. Chinese hamster ovary (CHO) cells, transfected with human insulin receptor constructs and expressing 10^6 receptors per cell (15), were grown in suspension in a 10-liter spinner flask in modified McCoy's 5A medium (GIBCO) containing 0.0345 mg of L-proline per ml and no $CaCl_2$ (16, 17). Cells were solubilized in 1.0% Triton X-100 and receptors were partially purified on wheat germ agglutinin (WGA)/agarose as described (17, 18). WGA-purified protein was stored at −70°C until needed. Identical aliquots of receptor were used for all assays.

Substrate Phosphorylation Experiments. WGA-purified insulin receptor (600 μl; 1.5 pmol) was incubated sequentially with 1 μM insulin (1 h at 4°C) and 50 μM ATP/5 mM $MnCl_2$ (1 h at 22°C) for maximal phosphorylation. The peptides were diluted appropriately in a mixture of 50 μM [γ-^{32}P]ATP/5 mM $MnCl_2$ in 50 mM Hepes containing 0.1% Triton X-100. Substrate phosphorylation reactions were initiated by the

Abbreviations: IRS-1, insulin receptor substrate 1; PDGF, platelet-derived growth factor; middle T antigen, polyoma virus middle-sized tumor antigen; PI 3-kinase, phosphatidylinositol 3′-kinase; SH2 and SH3, *src* homology domains 2 and 3; WGA, wheat germ agglutinin.

†To whom reprint requests should be addressed.

addition of 13 μl (32 fmol) of activated (insulin stimulated and ATP phosphorylated) insulin receptor, allowed to proceed at 22°C for 5 min (final vol, 40 μl), and terminated by addition of 65 μl of 5% trichloroacetic acid. Incorporated phosphate was determined by a modification of the phosphocellulose adsorption method (19). High molecular weight species were removed by precipitation after incubation in the presence of 0.16% bovine serum albumin for 30 min at 4°C. Portions of each supernatant solution containing phosphorylated peptides were spotted onto 2-cm^2 pieces of P81 phosphocellulose paper (Whatman) and washed four times for 30 min each in 1.0 liter of 0.075 M phosphoric acid. The papers were rinsed in acetone and allowed to air dry; incorporated phosphate was determined by Cerenkov counting.

Kinetic Analyses. Values for ^{32}P incorporation into the peptide substrates (cpm) were converted to pmol/min rates (V) to facilitate direct comparisons between peptides; dividing each pmol/min value by 1.3×10^{-5} mg (the amount of WGA-purified protein) converts it to the corresponding pmol per min per mg of protein value. K_m and V_{max} values and the corresponding standard errors were determined with the assistance of the ENZYME program (20), which follows algorithms for appropriate weighting of data described by Cleland (21); each peptide was analyzed in two to five separate experiments (see Table 2). The amount of insulin receptor in the WGA-purified protein isolated from human insulin receptor construct-transfected CHO cells, estimated by [^{125}I]insulin binding and Scatchard analyses using the LIGAND program (22), was 25 nmol/mg (range, 10–40 nmol/mg), which corresponds to 1% of the lectin-bound protein. This value was used to determine turnover rates (k_{cat}) and catalytic efficiency (k_{cat}/K_m).

RESULTS

Peptide Design. Peptides corresponding to sequences of IRS-1 surrounding tyrosine residues were synthesized. All consensus sites for tyrosine phosphorylation within YMXM motifs were prepared, as were additional consensus sequences not found in YMXM motifs (Table 1). Each peptide was designed to include (*i*) the specificity residue Y, (*ii*) acidic residues N-terminal to Y, (*iii*) residues at Y + 1 to Y + 5 positions, including the entire YMXMSP sequence, when appropriate, and (*iv*) as many native basic residues as possible, up to a total of three. If sufficient basic residues were not present within the native sequence, lysine residues were added at the C or N termini to a total of three to guarantee adsorption to the phosphocellulose paper used in the phosphopeptide assay (23). Therefore, peptides varied in length (10–18 residues) and position of the tyrosine residue.

Each peptide contains only one tyrosine with the exception of Y46,[‡] which contains two side-by-side tyrosines flanked by glutamic acid residues (Table 1). Peptides Y460 and Y546 contain YXXM motifs in which the Y + 1 position is isoleucine or threonine, respectively, with acidic residues at appropriate positions relative to Y (Table 1). Peptides Y608, Y628, Y658, Y727, Y939, and Y987 each contain complete YMXM motifs with at least one acidic residue N-terminal to Y (Table 1). A final peptide, Y998, contains a single tyrosine residue that is neither in a YMXM motif nor adjacent to an N-terminal acidic residue; like four of the YMXM peptides, however, it does contain a serine and proline at the Y + 4 and Y + 5 positions (Table 1).

[‡]Peptides are named according to the position of tyrosine in the sequence of IRS-1 (e.g., a peptide corresponding to the sequence surrounding Y^{987} is called Y987); an analogue of Y^{987} in which methionine at the +1 position is mutated to threonine is called Y987(M988T). Peptide sequences are listed in Table 1.

Table 1. Synthetic peptides corresponding to native and modified sequences of IRS-1

Position/name	Sequence
Double tyrosine	
Y46	RLE**Y**Y**E**NEKK
YXXM motifs	
Y460	*K*RGEEELSN**Y**I**C**MGGK
Y546	*KK*VSIEE**Y**T**E**MMPA*K*
YMXM motifs	
Y608	*KK*HTDDG**Y**M**P**MSPGVA
Y628	RKGNGDG**Y**M**P**MSPKSV
Y658	*KK*RVDPNG**Y**M**M**MSPSGS
Y727	*KKK*LPATGD**Y**M**N**MSPVGD
Y939	*KK*GSEE**Y**M**N**MDLGPGR
Y987	*KK*SRGD**Y**M**T**MQIG
Nonspecific sequence	
Y998	*K*PRNS**Y**VDT**S**PVAP*K*
Modified Y987 sequences	
Y987(D986N)	*KK*SRG*N***Y**M**T**MQIG
Y987(M988I)	*KK*SRGD**Y**I**T**MQIG
Y987(M988T)	*KK*SRGD**Y**T**T**MQIG
Y987(M988Nle)	*KK*SRGD**Y**†**T**MQIG
Y987(M990T)	*KK*SRGD**Y**M**T**TQIG

Phosphorylated tyrosines are boldface and underlined, as are additional elements of the YMXM specificity motif. Non-native residues are in italics; basic residues were added for assay purposes and there is a cysteine to alanine substitution for peptide Y727. For modified Y987 sequences, substituted positions are italicized.
†Norleucine.

Additional analogues of Y987 (a low K_m substrate with a complete YMXM motif) were prepared to test the functional importance of residues surrounding the targeted tyrosine (Table 1). To determine directly the effect of isoleucine or threonine in place of methionine at the Y + 1 position, as occurs naturally in sequences surrounding Y460 and Y546, these residues were substituted into the sequence of Y987. In addition, norleucine, which is isomorphic with methionine, was substituted at the Y + 1 position to further test requirements for side-chain hydrophobicity and flexibility. The function of methionine at the Y + 3 position of Y987 was also tested by analogous substitution to threonine, and the requirement for an acidic residue (aspartic acid at Y − 1) was tested by a conservative substitution to asparagine.

Peptide Phosphorylations and Kinetic Analyses. All peptides (Table 1) were phosphorylated by the insulin-stimulated and autoactivated receptor kinase in a time- and temperature-dependent fashion and displayed saturable kinetics (Fig. 1 *A* and *C*), which yielded linear Lineweaver–Burk plots (Fig. 1 *B* and *D*). In every case, phosphorylation reactions were linear for at least 5 min (data not shown), validating conditions used for determinations of K_m and V_{max} values (Table 2). Rates of peptide phosphorylation (V_{max}) and peptide concentrations required for half-maximal saturation (K_m) varied dramatically between peptides (Fig. 1). The slopes of double reciprocal plots, which are proportional to K_m/V_{max} (the inverse of catalytic efficiency), clustered into two groups (Fig. 1 *B* and *D*). Notably, native sequences having entire YMXM motifs displayed the smallest slopes (Fig. 1*D*) and were therefore phosphorylated most efficiently. By contrast, YXXM peptides and other sequences not containing complete YMXM motifs all displayed steeper slopes (Fig. 1*B*), demonstrating that they were phosphorylated less efficiently.

Calculation of kinetic constants revealed that K_m values for YMXM peptides ranged from 24 to 92 μM, lower than values previously reported for any peptides used as exogenous substrates of the insulin receptor (Table 2). Values for V_{max}, which range from 0.9 to 1.7 pmol/min (69–131 nmol·min^{-1}·mg^{-1}), are difficult to compare to previous studies as

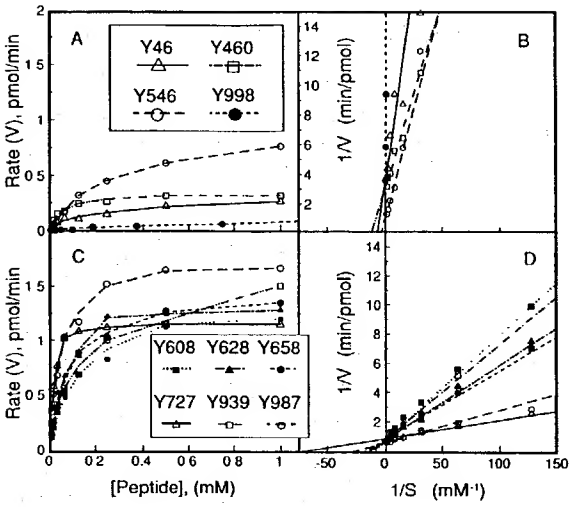

FIG. 1. Phosphorylation of peptides corresponding to native IRS-1 sequences by the insulin receptor kinase. Representative V vs. substrate concentration (A and C) and double reciprocal plots (B and D) are shown for phosphorylation of non-YMXM sequences (A and B) and YMXM sequences (C and D). Peptide sequences are identified in Table 1; methods used for phosphorylation reactions and data handling are described in *Experimental Procedures*.

this number varies with kinase specific activity. Values for k_{cat}/K_m, determined in this study, range from 0.6 to 2.1 × 10^4 $M^{-1}\cdot s^{-1}$. Therefore, in terms of substrate binding (estimated by K_m), turnover rates (k_{cat}), and overall catalytic efficiency (k_{cat}/K_m), all of the YMXM peptide sequences are excellent substrates of the insulin receptor kinase.

Similar analyses were performed with YXXM motifs (peptides Y460 and Y546; Table 1). Interestingly, K_m for peptide Y460 phosphorylation (isoleucine at the Y + 1 position) was 73 μM, within the range of values for YMXM peptides; this peptide was a less efficient substrate because V_{max} was 0.30 pmol/min, well below the range for YMXM peptides, which suggests that side-chain flexibility in addition to hydrophobicity at the Y + 1 position might be necessary for high turnover and catalytic efficiency. By contrast, the reduced catalytic efficiency of peptide Y546 (threonine at the Y + 1 position) was due to a high K_m for phosphorylation, which at 300 μM was substantially greater than corresponding values for YMXM peptides; in this case V_{max}, at 1.4 pmol/min, was similar to the YMXM peptides. Within a limited context, these results begin to suggest that the unique character of methionine (being flexible and hydrophobic) performs a special function in enhanced catalytic efficiency.

Table 2. Summary of kinetic constants

Peptide	K_m, μM	V_{max}, pmol/min	k_{cat}, s^{-1}	k_{cat}/K_m, $M^{-1}\cdot s^{-1}$
Double tyrosine				
Y46	140 ± 85	0.3 ± 0.06	0.14	1.0 × 10^3
YXXM motifs				
Y460	73 ± 15	0.3 ± 0.03	0.16	2.1 × 10^3
Y546	300 ± 30	1.4 ± 0.07	0.73	2.4 × 10^3
YMXM motifs				
Y608	90 ± 20	1.1 ± 0.09	0.57	6.2 × 10^3
Y628	60 ± 8	1.2 ± 0.06	0.63	1.0 × 10^4
Y658	86 ± 16	1.6 ± 0.12	0.83	9.7 × 10^3
Y727	24 ± 2.9	0.9 ± 0.06	0.47	2.0 × 10^4
Y939	61 ± 29	1.7 ± 0.3	0.89	1.5 × 10^4
Y987	34 ± 8.8	1.4 ± 0.11	0.73	2.1 × 10^4
Nonspecific sequence				
Y998	3200 ± 600	0.4 ± 0.04	0.21	6.5 × 10
Modified Y987 sequences				
Y987(D986N)	250 ± 69	3.9 ± 0.6	2.0	8.1 × 10^3
Y987(M988I)	80 ± 15	0.9 ± 0.07	0.47	5.9 × 10^3
Y987(M988T)	250 ± 80	2.7 ± 0.44	1.4	5.6 × 10^3
Y987(M988Nle)	24 ± 7.1	1.3 ± 0.12	0.68	2.8 × 10^4
Y987(M990T)	370 ± 110	1.2 ± 0.13	0.63	1.7 × 10^3

To test this more directly, the Y + 1 position within YMXM peptide Y987 was substituted with isoleucine, threonine, and norleucine (Table 1). In fact, both M988I and M988T substitutions reduced catalytic efficiency (k_{cat}/K_m) nearly 4-fold (Fig. 2B, compare slopes); the reasons for reduced efficiency were different for the two peptides, however. The M988I substitution increased K_m 2- to 3-fold and reduced V_{max} nearly 2-fold (Table 2). By contrast, the M988T substitution increased both K_m (7- to 8-fold) and V_{max} (2-fold). Thus, in a single, defined YMXM sequence M988I and M988T substitutions have the same general effect as observed for the YXXM peptides (Y460 and Y546). Substitution of methionine with norleucine, whose side chain mimics that of methionine regarding both hydrophobicity and flexibility, had no observable effect on K_m, V_{max}, and k_{cat}/K_m (Fig. 2B; Table 2).

For further comparison, we substituted the Y + 3 methionine of peptide Y987 and the acidic residue N-terminal to tyrosine (Y − 1). Surprisingly, the M991T substitution had an effect even greater than the related Y + 1 substitutions (Fig. 2D), with a 12-fold reduction in catalytic efficiency (k_{cat}/K_m) resulting exclusively from an increase in K_m (Table 2). Therefore, methionine residues at Y + 1 and Y + 3 positions play a very special role in directing efficient catalysis by the insulin receptor kinase. Acidic residues, which exist near autophosphorylated tyrosines in the insulin receptor and other protein tyrosine kinases, are presumed to be important for intermolecular substrate recognition as well (23–27). The only acidic residue in the Y987 sequence was substituted with

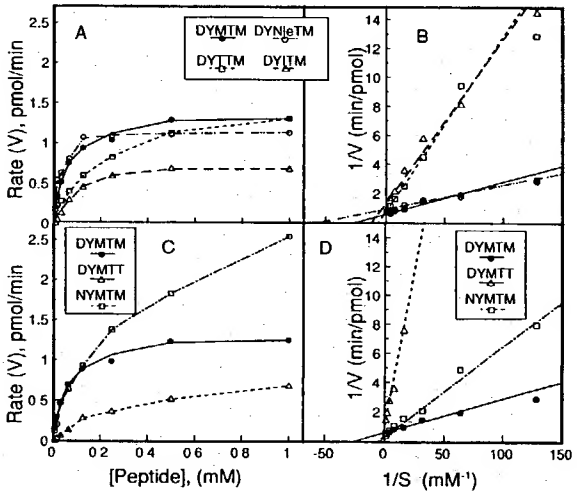

FIG. 2. Insulin receptor-catalyzed phosphorylation of substituted peptides corresponding to the IRS-1 sequence surrounding Y987. Representative V vs. substrate concentration (A and C) and double reciprocal plots (B and D) are shown for phosphorylation of peptides substituted at the Y + 1 position (A and B) or Y − 1 and Y + 3 positions (C and D). Peptide sequences are identified in Table 1; methods used for phosphorylation reactions and data handling are described in *Experimental Procedures*.

asparagine to assess the isolated effect of having no negative charge N-terminal to tyrosine (Fig. 2*D*). K_m was increased 8-fold while V_{max} was increased nearly 3-fold. These changes caused a net 2.5-fold reduction in catalytic efficiency (k_{cat}/K_m). Surprisingly, this is less deleterious than the corresponding effects observed for substituting methionine residues at either Y + 1 or Y + 3 positions (Table 2).

Two additional peptides corresponding to IRS-1 sequences not associated with YXXM or YMXM motifs were studied for comparison. Y46 contains a paired YY sequence flanked by acidic residues (EYYENE), resembling to a degree the major phosphorylation site of the insulin receptor itself (DIYETDYY) (28–30). The K_m for peptide Y46 phosphorylation was 134 μM, higher than that observed for YMXM peptides (Table 2), and V_{max} was 3- to 4-fold lower. Therefore, although Y46 can be considered to be a good substrate of the insulin receptor, it was phosphorylated with lower efficiency (6- to 36-fold) than YMXM peptides. Peptide Y998, which contains serine and proline at the Y + 4 and Y + 5 positions but lacks an acidic residue and methionine altogether, has a K_m for phosphorylation that is markedly elevated (3.2 mM). In addition, V_{max} is reduced (0.4 pmol/min) so k_{cat}/K_m is dramatically reduced (65 $M^{-1}\cdot s^{-1}$), which demonstrates the importance of primary sequence in directing insulin receptor kinase action and suggests that the effects of acidic residues N-terminal to tyrosine and methionine residues at Y + 1 and Y + 3 positions are cumulative.

DISCUSSION

Unlike serine/threonine-protein kinases, whose substrate recognition sequences are well defined by basic residues neighboring the phosphate acceptor (31–33), recognition sequences for substrates of protein tyrosine kinases are not well characterized. Autophosphorylated tyrosines within protein tyrosine kinase sequences are frequently located near acidic residues, which has led most investigators to focus on these positions as consensus sequences for tyrosine phosphorylation (23–27, 34–37). Synthetic peptides corresponding to autophosphorylation sites are phosphorylated by tyrosine kinases with K_m values typically in the millimolar range, which is considerably higher than values obtained for peptide substrates of serine/threonine kinases (23–27, 31–37). Unrelated peptides such as angiotensin and gastrin analogues are also phosphorylated by protein tyrosine kinases. Angiotensins I and II (both containing the DRVY*IHPF sequence) are phosphorylated with K_m values in the 1–4 mM range (26, 34), while gastrin analogues having five sequential glutamates N-terminal to tyrosine are phosphorylated with K_m values in the 50–200 μM range (35, 36). In fact, of all synthetic peptides studied previously as tyrosine kinase substrates, gastrin had the lowest K_m, a fact supporting the notion that acidic residues are important for tyrosine phosphorylation. However, systematic studies to define substrate specificity showed that when acidic residues of various peptides were replaced by uncharged amino acids, effects on V_{max}/K_m were often relatively small (2- to 4-fold) (23–27, 34–37).

The insulin receptor has tyrosine kinase activity similar to that of the viral transforming proteins, their cellular counterparts, and growth factor receptors. The insulin receptor phosphorylates the src peptide and angiotensin with K_m values in the millimolar range (26) and, in addition, phosphorylates a peptide corresponding to its own major autophosphorylation site (residues 1154–1164) (28–30) with a K_m value of 0.2–0.3 mM (18, 23, 30, 37). While the sequence at residues 1154–1164 might be a preferred substrate sequence, it was the recent identification of an endogenous substrate that allowed us to test sequences that might direct intermolecular substrate recognition. Finding biologically important sites of substrate phosphorylation will facilitate testing the importance of elements of secondary or tertiary structure in addition to that of primary sequence. For the insulin receptor, IRS-1 may be such a substrate. IRS-1 is a component of pp185 that is immunoprecipitated from insulin-stimulated cells by anti-phosphotyrosine antibodies. IRS-1 is phosphorylated on tyrosine residues in response to insulin (5–7) and appears to link the receptor to other components of the intracellular signaling cascade (e.g., PI 3-kinase; ref. 7).

As yet, we do not know which of the 34 tyrosine residues in IRS-1 are actually phosphorylated, although 14 have acidic residues N-terminal to the tyrosine position, which would categorize them as conventional consensus sequences for tyrosine kinase recognition. Surprisingly, 6 of these are within YMXM motifs and 2 more are in YXXM sequences, which previous studies (8–10, 38) suggest are important for interactions with PI 3-kinase and other proteins having SH2 domains. In this report, we show that synthetic peptides corresponding to IRS-1 YMXM sequences are excellent substrates for the insulin receptor kinase. In fact, calculated values of K_m are lower than has been reported previously for peptides phosphorylated by any tyrosine kinase, with the exception of epidermal growth factor receptor-catalyzed phosphorylation of gastrin (which contains the homologous EEEEEAYGWM sequence; refs. 35 and 36).

Amino acid substitution studies show that different positions within the YMXM motif effect catalytic efficiency. Substitutions of methionine at Y + 1 decrease catalytic efficiency for tyrosine phosphorylation ≈4-fold, estimated by (*i*) comparisons of k_{cat}/K_m for peptides Y460 and Y546 (YXXM motifs) and the six YMXM peptides, and (*ii*) direct comparisons between k_{cat}/K_m for the wild-type Y987 sequence and substituted peptides Y987(M988I) and Y987(M988T). Notably, norleucine and methionine at the Y + 1 position are functionally indistinguishable. Substitution of methionine at Y + 3 with threonine has an even greater effect on peptide Y987 phosphorylation, in this case reducing k_{cat}/K_m ≈12-fold due to an effect on K_m. Both of these effects are greater than that produced by substituting aspartic acid with asparagine at Y − 1 of peptide Y987 (≈2.5-fold reduction in k_{cat}/K_m). Sequences having neither N-terminal acidic residues nor methionine at Y + 1 and Y + 3 positions are phosphorylated much less efficiently. We do not know the significance of residues at the Y + 2 position (proline, methionine, asparagine, or threonine in the YMXM sequences studied here). Four of the six YMXM sequences of IRS-1 also have serine and proline at the Y + 4 and Y + 5 positions, although peptides with or without the YMXMSP residues are phosphorylated with equal efficiency. Furthermore, the Y998 sequence contains serine and proline at the Y + 4 and Y + 5 positions but without an acidic residue or methionine phosphorylation was inefficient. The function (if any) served by these residues appears not to be related to phosphorylation.

What, if anything, is special about Y + 1 and Y + 3 positions, and why does methionine at these positions enhance kinetic efficiency? Sequences directly contiguous to and including YMXM are predicted by the Chou–Fasman algorithm (41) to be unstructured, although β-turns are predicted to flank each of the YMXM motifs of IRS-1 (data not shown). Alternatively, it has been proposed that substrates might adopt amphiphilic helical structures at the kinase catalytic surface (42); if so, methionine residues at Y + 1 and Y + 3 positions might be positioned appropriately for direct participation in binding. While we have not yet been able to analyze structure, the unique character of methionine may play a special role here. Isoleucine, leucine, and valine, while also highly hydrophobic, have branched and thus relatively rigid side chains. By contrast, the side chain of methionine is unbranched, providing considerable structural flexibility. In support of this, norleucine, which is equally flexible and hydrophobic compared to methionine, has an identical capacity to direct insulin receptor kinase action. This feature of methionine has recently

been proposed to provide a malleable nonpolar surface for signal peptide recognition (43) and plasticity in protein interactions with calmodulin (39), which would allow structurally diverse nonpolar surfaces to conform to the respective recognition surfaces of the proteins to which they bind (40). Such a malleable, nonpolar surface might provide a mechanism to explain how the same IRS-1 YMXM motifs interact with the insulin receptor kinase and PI 3-kinase and, potentially, other tyrosine kinases and proteins with SH2 domains as well. The side chains of residues in the YMXM domain might be sufficiently flexible to provide a nonpolar surface that accommodates itself equally to the surfaces of kinase active sites and SH2 protein binding pockets.

Relationships between IRS-1 and other signal transduction proteins may be quite complex, as IRS-1 contains multiple copies of the YMXM motif that could interact selectively with different tyrosine kinases as well as distinct effector molecules containing various isoforms of the SH2/SH3 domain. Furthermore, other proteins involved in signal transduction contain related sequences. The PDGF receptor phosphorylates itself within YMXM or YVXM motifs (44), the src-associated middle T antigen contains a phosphorylated EEEEEEYMPM sequence (8, 45), and similar sequences within the colony-stimulating factor 1 receptor (DTYVEM) and *kit* oncogene (DSTNEYMDM) may be phosphorylated, as well (8). Whether YMXM motifs define a recognition sequence for efficient tyrosine phosphorylation by these and other protein tyrosine kinases remains to be tested. For the src-associated middle T antigen and the PDGF receptor, these sequences are thought to be involved in recognition by PI 3-kinase through interaction with one or the other of its SH2 domains (8, 9, 38, 44). Similarly, studies to block interactions between IRS-1 and PI 3-kinase with synthetic phosphopeptides corresponding to IRS-1 YMXM sequences suggest that these interactions involve SH2 domains as well (unpublished data).

In conclusion, YMXM, a redundant sequence motif identified in IRS-1, is shown to define substrate specificity of the insulin receptor kinase. Methionine residues at the Y + 1 and Y + 3 positions are particularly important for efficient kinase recognition, in addition to acidic residues N-terminal to tyrosine. Knowing a target motif for the action of the insulin receptor kinases, and possibly other tyrosine kinases as well, may enhance our understanding of tyrosine kinase signaling pathways. Furthermore, a tyrosine kinase target motif may facilitate identification of additional known or newly identified proteins that might be involved in related signaling pathways. Certainly, as genomic sequencing efforts progress, it will be increasingly useful to have identified short peptide motifs to help decipher which of these DNA sequences encodes proteins acting as potential targets of tyrosine kinase action.

This work was supported in part by the National Science Foundation (S.E.S.), National Institutes of Health (M.F.W.), and Diabetes and Endocrinology Research Center Grant DK08366. S.C. is supported by a Fellowship from the Juvenile Diabetes Foundation; S.E.S. is a Capps' Scholar in Diabetes at Harvard Medical School and the recipient of a Career Development Award from the Juvenile Diabetes Foundation.

1. Kasuga, M., Zick, Y., Blithe, D. L., Crettaz, M. & Kahn, C. R. (1982) *Nature (London)* **298**, 667–669.
2. Kasuga, M., Karlsson, F. A. & Kahn, C. R. (1982) *Science* **215**, 185–187.
3. Rosen, O. M., Herrera, R., Olowe, Y., Petruzzelli, L. M. & Cobb, M. H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 3237–3240.
4. Yu, K. & Czech, M. P. (1984) *J. Biol. Chem.* **259**, 5277–5286.
5. White, M. F., Maron, R. & Kahn, C. R. (1985) *Nature (London)* **318**, 183–186.
6. Rothenberg, P. L., Lane, W. S., Karasik, A., Backer, J. M., White, M. F. & Kahn, C. R. (1991) *J. Biol. Chem.* **266**, 8302–8311.
7. Sun, X. J., Rothenberg, P., Kahn, C. R., Backer, J. M., Araki, E., Wilden, P. A., Cahill, D. A., Goldstein, B. J. & White, M. F. (1991) *Nature (London)* **352**, 73–77.
8. Cantley, L. C., Auger, K. R., Carpenter, C., Duckworth, B., Graziani, A., Kapeller, R. & Soltoff, S. (1991) *Cell* **64**, 281–302.
9. Escobedo, J. A., Kaplan, D. R., Kavanaugh, W. M., Turck, C. W. & Williams, L. T. (1991) *Mol. Cell. Biol.* **11**, 1125–1132.
10. Koch, C. A., Anderson, D., Moran, M. F., Ellis, C. & Pawson, T. (1991) *Science* **252**, 668–674.
11. Moran, M. F., Koch, C. A., Anderson, D., Ellis, C., England, L., Martin, G. S. & Pawson, T. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 8622–8626.
12. Ruderman, N. B., Kapeller, R., White, M. F. & Cantley, L. C. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 1411–1415.
13. Edemann, G., Yohezawa, K. & Roth, R. A. (1990) *J. Biol. Chem.* **265**, 396–400.
14. Tam, J. P. & Merrifield, R. B. (1987) in *The Peptides: Analysis, Synthesis, Biology*, eds. Udenfriend, S. & Meienhofer, J. (Academic, New York), pp. 185–248.
15. Ebina, Y., Edery, M., Ellis, L., Beaudoin, J., Roth, R. A. & Rutter, W. J. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 8014–8018.
16. Shymko, R. M., Gonzales, N. S., Backer, J. M., White, M. F. & De Meyts, P. (1989) *Biochem. Biophys. Res. Commun.* **164**, 191–198.
17. Shoelson, S. E., Lu, Z., Parlautan, L., Lynch, C. S. & Weiss, M. A. (1992) *Biochemistry*, in press.
18. Shoelson, S. E., White, M. F. & Kahn, C. R. (1988) *J. Biol. Chem.* **263**, 4852–4860.
19. Glass, D. B., Masaracchia, R. A., Feramisco, J. R. & Kemp, B. E. (1978) *Biochem. J.* **87**, 566–575.
20. Lutz, R. A., Bull, C. & Rodbard, D. (1986) *Enzyme* **36**, 197–206.
21. Cleland, W. W. (1979) *Methods Enzymol.* **63**, 103–138.
22. Munson, P. J. & Rodbard, D. (1980) *Anal. Biochem.* **107**, 220–239.
23. Stadtmauer, L. A. & Rosen, O. M. (1986) *J. Biol. Chem.* **261**, 10000–10005.
24. Hunter, T. (1982) *J. Biol. Chem.* **257**, 4843–4848.
25. Patschinsky, T., Hunter, T., Esch, F. S., Cooper, J. A. & Sefton, B. M. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 973–977.
26. Stadtmauer, L. A. & Rosen, O. M. (1983) *J. Biol. Chem.* **258**, 6682–6685.
27. Geahlen, R. L. & Harrison, M. L. (1989) in *Peptides and Protein Phosphorylation*, ed. Kemp, B. E. (CRC, Boca Raton, FL), pp. 239–250.
28. Ebina, Y., Ellis, L., Jarnagin, K., Edery, M., Graf, L., Clauser, E., Ou, J.-H., Masiarz, F., Kan, Y. W., Goldfine, I. D., Roth, R. A., & Rutter, W. J. (1985) *Cell* **40**, 747–758.
29. Ullrich, A., Bell, J. R., Chen, E. Y., Herrera, R., Petruzzelli, L. M., Dull, T. J., Gray, A., Coussens, L., Liao, Y.-C., Tsubokawa, M., Mason, A., Seeburg, P. H., Grunfeld, C., Rosen, O. M. & Ramachandran, J. (1985) *Nature (London)* **313**, 756–761.
30. White, M. F., Shoelson, S. E., Keutmann, H. & Kahn, C. R. (1988) *J. Biol. Chem.* **263**, 2969–2980.
31. Kemp, B. E. & Pearson, R. B. (1990) *Trends Biochem. Sci.* **15**, 342–346.
32. Kemp, B. E., ed. (1990) *Peptides and Protein Phosphorylation* (CRC, Boca Raton, FL).
33. Kennely, P. J. & Krebs, E. G. (1991) *J. Biol. Chem.* **266**, 15555–15558.
34. Wong, T. W. & Goldberg, A. R. (1983) *J. Biol. Chem.* **258**, 1022–1025.
35. Baldwin, G. S., Burgess, A. W. & Kemp, B. E. (1982) *Biochem. Biophys. Res. Commun.* **109**, 656–662.
36. Baldwin, G. S., Knesel, J. & Monckton, J. M. (1983) *Nature (London)* **301**, 435–438.
37. Shoelson, S. E., White, M. F. & Kahn, C. R. (1989) *J. Biol. Chem.* **264**, 7831–7836.
38. Auger, K. R., Carpenter, C. L., Shoelson, S. E., Piwnica-Worms, H. & Cantley, L. C. (1992) *J. Biol. Chem.*, in press.
39. O'Neil, K. T. & DeGrado, W. F. (1990) *Trends Biochem. Sci.* **15**, 59–64.
40. Gellman, S. H. (1991) *Biochemistry* **30**, 6633–6636.
41. Fasman, G., ed. (1989) *Prediction of Protein Structure and the Principles of Protein Conformation* (Plenum, New York).
42. Radziejewski, C., Miller, W. T., Mobashery, S., Goldberg, A. R. & Kaiser, E. T. (1989) *Biochemistry* **28**, 9047–9052.
43. Bernstein, H. D., Poritz, M. A., Strub, K., Hoben, P. J., Brenner, S. & Walter, P. (1989) *Nature (London)* **340**, 482–486.
44. Escobedo, J. A., Navankasattusas, S., Kavanaugh, W. M., Milfay, D., Fried, V. A. & Williams, L. T. (1991) *Cell* **65**, 75–82.
45. Hunter, T., Hutchinson, M. A. & Eckhart, W. (1984) *EMBO J.* **3**, 73–79.